Bias and Efficiency for SEM with Missing Data and Auxiliary Variables: Robust Method versus Normal Distribution Based ML

Ke-Hai Yuan*, Xin Tong, and Zhiyong Zhang

University of Notre Dame, USA (kyuan@nd.edu)
University of Notre Dame, USA
University of Notre Dame, USA

Abstract:

It is well-known that maximum likelihood estimates (MLEs) are inconsistent when missing values are missing not at random (MNAR) and the mechanism is ignored in the estimation process, and including proper auxiliary variables can change missing data mechanism to missing at random (MAR) and results in consistent estimates. However, when including auxiliary variables that are of heavier tails, the widely used normal-distribution-based MLEs (NMLEs) may have much greater variances. Recent results indicate that auxiliary variables that are nonlinear functions of the observed variables are also needed for NMLEs to be consistent even when missing values are MAR, and again, the NMLEs may have much greater variances. Conceivably, robust methods will generate more precise estimates than NMLEs in both cases. This study compares the NMLEs against robust estimates when including different type of auxiliary variables: Some are observed variables and some are quadratic or cubic functions of the observed variables. The effect of auxiliary on the change of parameter estimates is also studied to gauge possible MNAR mechanism or nonlinear relationships among the observed variables. In the context of a confirmatory factor model, missing data mechanisms are created so that the missing probabilities are functions of either the observable variables or the common factor or errors/uniqueness.